Dynamics of reward based decision making a computational study

Bhargav Teja Nallapu[1,2] and Nicolas P. Rougier[2,3,4]

[1] International Institute of Information Technology (I.I.I.T), Hyderabad
[2] INRIA, Bordeaux Sud-Ouest, Talence, France
[3] IMN, CNRS, University of Bordeaux, UMR 5293, IMN, Bordeaux, France
[4] University of Bordeaux, CNRS UMR 5800, Labri, IPB, Talence, France

Abstract. We consider a biologically plausible model of the basal ganglia that is able to learn a probabilistic two armed bandit task using reinforcement learning. This model is able to choose the best option and to reach optimal performances after only a few trials. However, we show in this study that the influence of exogenous factors such as stimuli salience and/or timing seems to prevail over optimal decision making, hence questioning the very definition of action-selection. What are the ecological conditions for optimal action selection ?

Keywords: Decision making, neural dynamics, basal ganglia, optimal behavior

1 Introduction

Basal ganglia are known to be involved in decision making and action selection based on reinforcement learning and a number of models have been designed to give account on such action selection [10, 2, 3]. We have been studying a specific computational model of the basal ganglia that has been introduced in [4] and replicated in [13]. This model has been used to explain, to some extent, decision making in primates on a two armed bandit task. One of the questions we attempt to address in this study is to what extent the physical properties of the stimulus such as the visual salience or other characteristics affect the decision and lead to a suboptimal choice. For example (and quite obviously), a stimulus is very likely to be selected, if it is presented before the other stimuli and this selection will be made irrespectively of the potential reward associated with this stimulus. Moreover, there may be other factors such as stimulus salience or population size that may also disrupt the optimal performance. This led us to do a systematic study of the influence of such exogenous factors to understand what are the ecological conditions for optimal decision making.

2 Methods

2.1 Task

The task that has been used to demonstrate action selection in the model is a probabilistic learning task that is described in [8]. Four target shapes are

associated with different reward probabilities (see figure 1). A trial is a time period in which any two of the four possible shapes are presented at two random positions (out of the four possible positions - up, right, down and left). The model is allowed to settle for the first 500ms of the trial and then two random cues are presented. By the end of trial period, a choice is made and the reward is given according to the reward probability associated with the chosen shape. A trial is

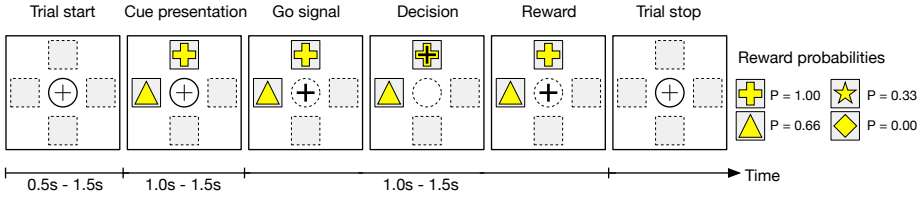

Fig. 1. The two armed bandit task as described in [8, 4].

considered to be successful if a decision is made by the model, irrespective of the reward received. In a single independent trial, the cognitive decision (shape of the cue) and motor decision (position of the cue) are independent of each other. At any decision-making level of the model, each of the four cue shapes and each of the four motor movement directions is represented by one unit (neuron) each. Thus in a given trial, when two cue shapes are presented at two different positions, two cognitive, two motor and two associative (in cortex and striatum, see figure 2) neurons are activated. The task is run for a session, a number of trials, while at the end of each trial, the model learns the reward associated to its selection. (see *Learning*).

2.2 Model

In [5], authors demonstrated an action selection mechanism in the cortico-basal ganglia loops based on a competition between the positive feedback, direct pathway through the striatum and the negative feedback, hyperdirect pathway through the subthalamic nucleus. In [4], authors investigated further how multiple level action selection could be performed by the basal ganglia, and the model has been extended in a manner consistent with known anatomy and electrophysiology of the basal ganglia in the monkey (see figure 2). This model allows a bidirectional information flow between loops such that during early trials, a direction can be selected randomly, irrespective of the cue positions. However, after repeated trials, the model is able to consistently make the cognitive decision before the motor decision in each trial (see figure 3) and most frequently the motor decision, biased by the cognitive decision, towards the position of the more rewarding cue shape.

Learning . Learning has been derived from a simple actor-critic algorithm [12] that shapes the gain between the cognitive cortex and the cognitive striatum.

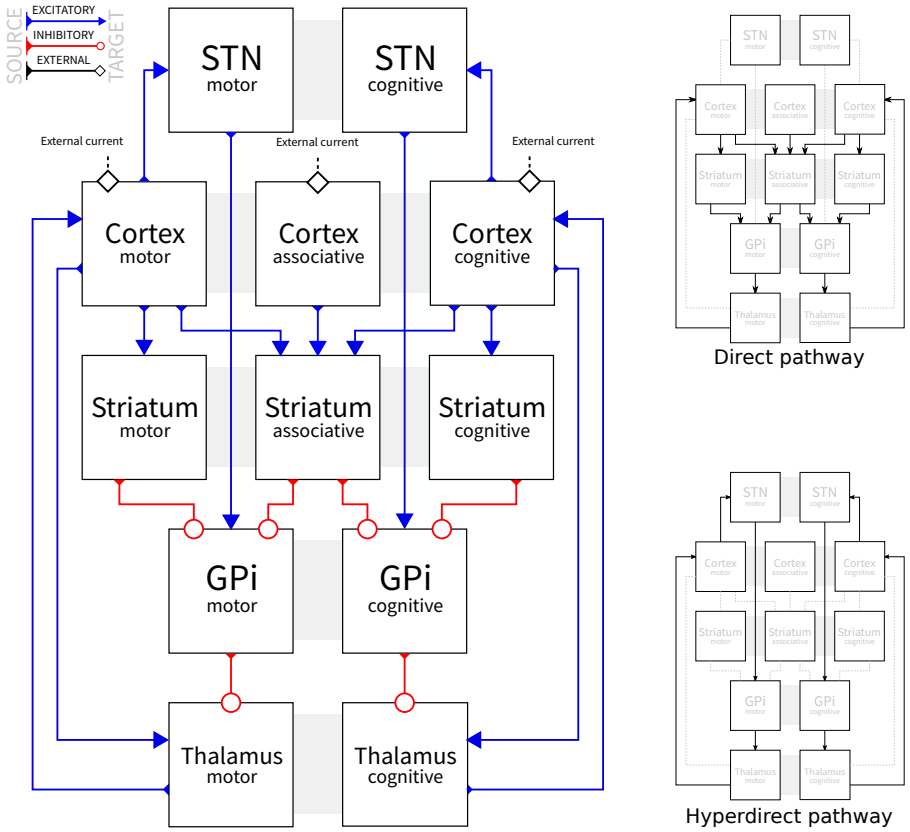

Fig. 2. Architecture of the basal ganglia model which is organised around three parallel loops: cognitive, associative and motor. Only the direct and hyperdirect pathways have been modelled.

According to the amount of reward received at the end of each trial (0 or 1), the model learns to estimate the value of chosen stimulus and then the cognitive pathway is biased in favor of the stimulus with the highest value.

Neuronal dynamics . This model uses the same, simple neuronal rate model as in [5] and [4] to focus on the network dynamics. Within each structure, each neuron (in each channel of any loop) is modeled as a single rate coded neuron with the equation:

$$\tau \frac{dm}{dt} = -m + I_s + I_{Ext} - T \tag{1}$$

decay time constant of the synaptic input τ, negative values of activation, m and threshold of the neuron T are set to respective constant values as per the model in [4]. I_{Ext} is an external input representing the sensory visual salience of the cue, which is unchanged throughout the process of learning.

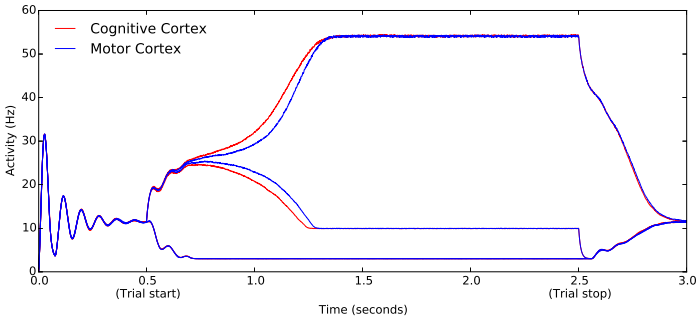

Fig. 3. Time course of a decision in the motor cortex (blue curves) and cognitive cortex (red curve) before learning. At trial start (t=500ms), there is a first bifurcation between stimuli that are actually presented and those who are not. The second bifurcation around t=750ms is the actual decision of the model.

3 Results

In all the following cases, we consider 4 stimuli A, B, C, D respectively associated with reward probability of 1.0, 0.66, 0.33 and 0.0. Learning is performed over 120 trials until the model reaches a performance of 0.90, meaning it chooses the best stimulus 90% of the time. We then stopped learning in the model and simulated a scenario where one stimulus is presented first and the other follows after a certain *delay*. Another scenario involved presenting one stimulus with more *salience* than the other. In both the scenarios, the intent was to emphasize the advantage (earlier presentation or higher salience) to the lesser rewarding stimulus and see if that leads the model to a suboptimal decision. We presented the model with various scenarios involving different *delays* and *saliences*. (see figure 4).

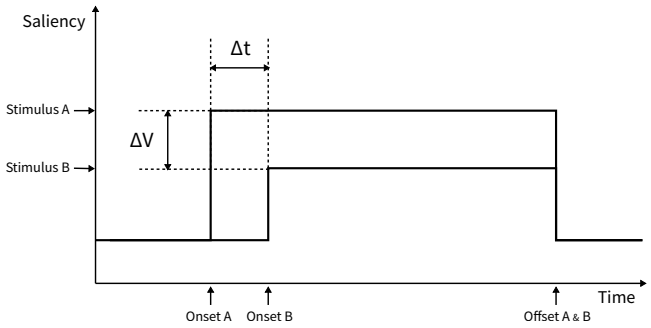

Fig. 4. Two stimuli A & B can differ in salience (ΔV) and/or in timing (Δt). ΔV is expressed as the relative ratio between the less salient and the most salient stimuli ($\Delta V = (V_A - V_B)/V_B$. Δt is expressed as the delay separating the two stimuli onsets ($\Delta t = t_A - t_B$).

3.1 Influence of delay

We first tested the influence of a small delay (between 0ms and 60ms) between the presentation of the two stimuli. The worst stimulus, that is the one associated with the lesser probability of reward, is presented first and after the delay, the second (better rewarding) stimulus is presented. We have been testing systematically all combinations of stimuli (A/B, A/C, A/D, B/C, B/D, C/D) and averaged the mean performance over 25 trials (see figure 5). As expected, the performance decreased with the increase in delay and the crossing (i.e. performance is random) happens around 35ms for all combinations but the last one (C/D) that happens very early, around 20ms. This specific case can be explained by the poor estimation of the value of C and D during learning because those stimuli are almost never chosen (most of the time, they are presented with a better stimuli).

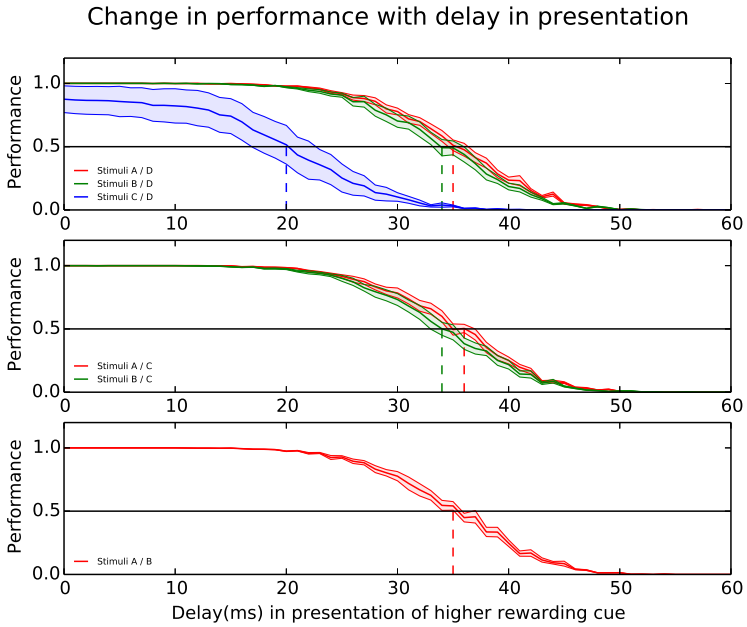

Fig. 5. Performance of the model as a function of the delay between the worst and the best stimuli. All combinations have been tested and mean performance has been averaged over 25 trials.

3.2 Influence of salience

We tested the influence of salience by presenting simultaneously the two stimuli but the worst stimulus, i.e, the one associated with a lesser probability of reward,

has been made virtually stronger than the other. The model, having learned the rewarding probabilities, is expected to select the higher rewarding stimulus irrespective of the salience. However, the increased salience of the lesser rewarding stimulus affects the model and leads it to take a bad decision (see figure 6). As the salience of the lesser rewarding stimulus increases, a consistent decrease in the performance of model is observed. Interestingly, the threshold percentage of salience difference after which the performance of the model decreases, is a characteristic of the difference in the reward probabilities of both the stimuli presented. Quite visibly (figure 6), it takes higher salience difference for a lesser rewarding stimulus to be chosen against the best rewarding one (in this case, A) whereas a lesser increase in salience seems to be sufficient to compromise the decisions involving lesser rewarding stimuli, like B.

In various neuropsychological studies on humans, like in [7] and [9], it has been emphasized that the visual saliency of stimuli influences the choices over the learned preferences and visual working memory. Interestingly, in [7] where at an exposure time of 1500ms, which is quite similar to that of the model discussed here, the influence of visual saliency was particularly evident when there were no strong preferences among the options. This observation is supported by the early performance decline of the model discussed here, when presented with two closely rewarding stimuli (Stimuli C/D in figure 6).

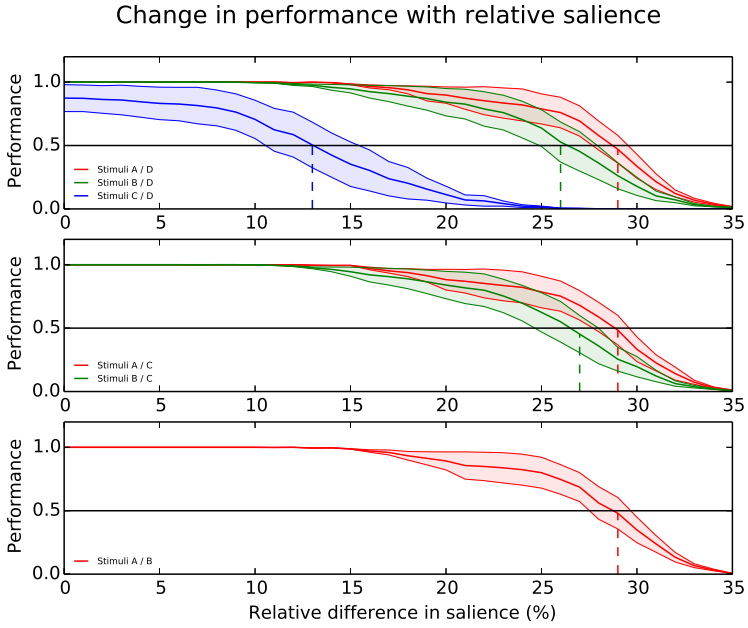

Fig. 6. Decrease in performance of the model when the lesser rewarding stimulus is presented with stronger salience than the higher rewarding stimulus.

3.3 Joint influence of delay and salience

We further tested the model and studied the joint influence of delay and salience by make the worst stimulus to be presented earlier and with a stronger salience. As shown in figure 7, this dramatically degrades the performance and the domain of optimal performance is even more restricted compared to original results.

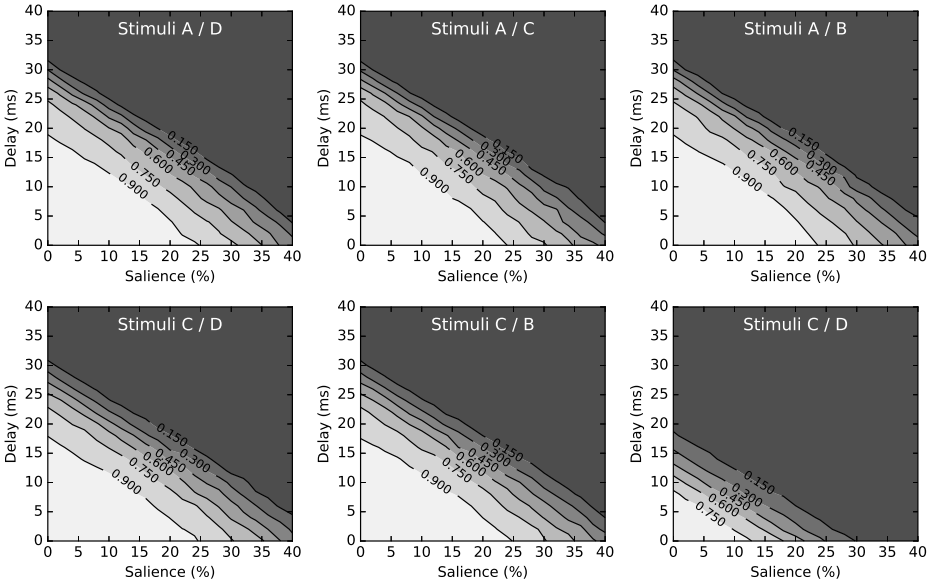

Fig. 7. Joint influence of delay and salience on the performance of the model. The two effects appears to linearly sum up and the domain of optimal performance is even more restricted.

4 Conclusion

These early results tend to question the very notion of optimal action selection as defined in a number of theoretical works. The action may be considered optimal provided the two options are presented simultaneously and with an equivalent representation. In the reinforcement learning paradigm, such consideration does not hold much relevance. However, from a more behavioral and embodied perspective, we think this is an important dimension to consider because an animal is scarcely confronted by a set of perfectly equivalent options (but their associated value). One may come first or one may just appear more "obvious" (i.e. more salient). In such a case, the inner dynamics of the model may lead to a suboptimal choice as it is the case using the model from [4]. Although we did not perform the study presented here on primates yet, the results from the model assert the need for a closer look at the way we perceive decision making paradigms.

The question is to know to what extent some dedicated brain mechanisms are able to cope with these problems. For example, concerning the time delay, a *stop* signal, as it has been reported in [11], may represent a potential mechanism to be able to solve the problem for small delays (< 200ms).

For the salience difference however, and to the best of our knowledge, there is no such *dedicated* mechanism. In [6], a stimulus-reward association study on macaque monkeys, spike recordings showed significant reward dependence in their responses to the visual cues. In [1], rewards were shown to teach visual selective attention maximizing the positive outcomes. However both the studies do not identify the underlying mechanisms that caused the observations on the effect of salience. This study hence suggests that measuring experimentally performance using different salience levels could bring useful insights into decision making.

References

1. Chelazzi, L., Perlato, A., Santandrea, E., Della Libera, C.: Rewards teach visual selective attention. Vision Research 85, 58–72 (2013)
2. Gurney, K., Prescott, T.J., Redgrave, P.: A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biological cybernetics 84(6), 401–410 (2001)
3. Gurney, K., Prescott, T.J., Redgrave, P.: A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. Biological cybernetics 84(6), 411–423 (2001)
4. Guthrie, M., Leblois, A., Garenne, A., Boraud, T.: Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. Journal of neurophysiology 109(12) (2013)
5. Leblois, A., Boraud, T., Meissner, W., Bergman, H., Hansel, D.: Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. Journal of Neurosciences 26, 3567–3583 (2006)
6. Mogami, T., Tanaka, K.: Reward association affects neuronal responses to visual stimuli in macaque te and perirhinal cortices. The Journal of neuroscience 26(25), 6761–6770 (2006)
7. Mormann, M.M., Navalpakkam, V., Koch, C., Rangel, A.: Relative visual saliency differences induce sizable bias in consumer choice. Journal of Consumer Psychology 22(1) (2012)
8. Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., Gross, C.E., Boraud, T.: Shaping of Motor Responses by Incentive Values through the Basal Ganglia. Journal of Neuroscience 27(5) (2007)
9. Pooresmaeili, A., Bach, D.R., Dolan, R.J.: The effect of visual salience on memory-based choices. Journal of neurophysiology 111(3), 481–487 (2014)
10. Redgrave, P., Prescott, T.J., Gurney, K.: The basal ganglia: a vertebrate solution to the selection problem? Neuroscience 89(4), 1009–1023 (1999)
11. Schmidt, R., Leventhal, D.K., Mallet, N., Chen, F., Berke, J.D.: Canceling actions involves a race between basal ganglia pathways. Nature Neuroscience (2013)
12. Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. MIT Press (1998)
13. Topalidou, M., Rougier, N.: [re] interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. ReScience 1(1) (2015)